

University of Groningen

## Sentence comprehension in monolingual and bilingual aphasia

Arantzeta Perez, Miren

**IMPORTANT NOTE:** You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

*Document Version*

Publisher's PDF, also known as Version of record

*Publication date:*

2017

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Arantzeta Perez, M. (2017). *Sentence comprehension in monolingual and bilingual aphasia: Evidence from behavioral and eye-tracking methods*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen.

### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

# CHAPTER 4

---

## What happens when they think they are right? Error awareness analysis of sentence comprehension deficits in aphasia

### Abstract

**Background:** Comprehension of non-canonical sentences is frequently characterized by chance-level performance in people with aphasia (PWA). Chance-level performance has been interpreted as guessing, but online data does not support this rendering. It is still not clear whether the incorrect sentence processing is guided by the compensatory strategies that PWA might employ to overcome linguistic difficulties. This study combined off-line and online data to investigate the effect of word order and error-awareness on sentence comprehension in a group of PWA and non-brain-damaged (NBD) speakers of Spanish. The off-line tasks involved auditory sentence-picture matching immediately followed by a confidence rating. Participants were asked to judge the perceived correctness of their previous answer. Online data consisted of eye-tracking. NBD participants showed sentence comprehension at ceiling level, while PWA showed comprehension difficulty, with the tendency to perceive as correct both correctly and incorrectly answered trials. Just 6.8% of judgments were classified as “guessing” by PWA. Confidence rating was a poor predictor of response accuracy in PWA, but moderate-good in NBD. Post-hoc gaze data analysis indicated that confidence rating was a predictor of the fixation pattern during the presentation of the linguistic stimuli. Results suggest that PWA were mostly unaware of their sentence comprehension errors and did not consciously employ strategies to compensate for their difficulties.

## 4.1 Introduction

### 4.1.1 Word order and sentence comprehension deficits in PWA

Sentence comprehension deficits in agrammatic aphasia have been well established in the literature over the past few decades. In the absence of lexical comprehension deficits, semantically reversible sentences presented in non-canonical<sup>13</sup> word orders are frequently misunderstood by people with non-fluent aphasia (PWA) (e.g., Caplan & Hildebrandt, 1988; Caramazza & Zurif, 1976; Grodzinsky, 2000; Thompson et al., 2013; Schumacher, et al., 2015; see Grodzinsky, Piñango, Zurif, & Draí, 1999 for a review). Both semantic reversibility and word order are key in understanding this difficulty. When both Determiner Phrases (DPs) of a sentence are animate and potential agents of the action (i.e., the sentence is reversible), PWA show better comprehension of sentences presented in Agent-Theme order (henceforth A-T) (e.g., *The girl calls the teacher*) than in Theme-Agent order (henceforth T-A) (e.g., *The teacher has been called by the girl*). This is because in the former, listeners may rely solely on word order information to disentangle the thematic-role assignment, while in the latter they are forced to process morphological information to eventually perform parsing routines and reach the correct interpretation of the sentence.

Typologically, the more (case and agreement) morphology a language has, the greater freedom it displays in sentence word order. Some authors have suggested that PWA are sensitive to language-specific cues and this premorbid awareness may impact their difficulties (Bates, Friederici, & Wulfeck, 1987). This suggests that the strength of word order as a reliable parsing cue varies cross-linguistically. That is, speakers of highly inflected languages may rely on morphological cues rather than word order information. Nevertheless, deficits in comprehending sentences in T-A argument order have been found in PWA across languages with more rigid word order (i.e., in English, Schwartz, Saffran, & Marin, 1980; Meyer, Mack, & Thompson, 2012; in Dutch; Bastiaanse & Edwards, 2004), flexible word order (i.e., in Spanish, Juncós-Rabadán, Pereiro, & Souto, 2009; in German, Burchert, De Bleser, & Sonntag, 2003; Hanne et al., 2011; in Italian, Garraffa & Grillo, 2008; in Turkish, Duman, Altınok, Özgirgin, & Bastiaanse, 2011; in Swahili, Abuom, Shah, & Bastiaanse, 2013; cf. in Indonesian, Jap, Martínez-Ferreiro, & Bastiaanse, 2016) and free word order (i.e., in Basque, Arantzeta et al., 2017, in Chapter 2). Nevertheless, a word order effect can be more prominent in some languages than in others (see Munarriz, Ezeizabarrena, & Gutierrez-Mangado, 2014; Vaid & Pandit, 1991), but it points out that

---

<sup>13</sup> In the current Chapter, the terms canonical vs. non-canonical refer to the base word order in a certain language (e.g., SVO in Spanish) vs. the rest of word orders allowed in the language (e.g., OVS, VSO, VOS). Aside from this, the terms linear vs. non-linear refer to the order of Agent and Theme in relation to each other; linear Agent-Theme (e.g., SVO, VSO) vs. non-linear Theme-Agent (e.g., OVS, VOS).

the processing of morphosyntactic information is a core deficit in agrammatic aphasia (see Thompson, Kiellar, & Fix, 2012 for an overview).

To date, most studies on sentence comprehension have used off-line methods to address the ways in which PWA and non-brain-damaged (NBD) individuals process sentences in order to assign grammatical functions and thematic roles. The sentence-picture matching task has been typically used in both experimental and clinical settings. The participant needs to choose, within a set of two (or more) visual stimuli, the one that best matches the target sentence. Although the results are easy to quantify, it is necessary to consider that the odds of picking the target picture by chance are relatively large. Thus, to compare results against chance, accuracy scores at chance-level (e.g., 33.3%–66.66% accuracy in a binomial choice task) have been traditionally attributed to guessing by the Trace Deletion Hypothesis (TDH; Grodzinsky, 1986, 1995, 2000; see Draai & Grodzinsky, 2006ab for a later revision). Based on the tenets of the Principles and Parameters model of generative grammar<sup>14</sup> (Chomsky, 1981), the TDH states that agrammatism precludes the creation of a chain between the moved element and the trace in its original position. According to the TDH, traces of movement are not available to PWA. Consequently, when presented with passive sentences with non-linear A-T order, PWA cannot assign thematic roles to a moved argument and apply instead a default strategy that assigns the role of Agent to the first DP in the sentence. The argument in the by-phrase gets the correct thematic role, and therefore, the sentence appears to have two agents. Accordingly, the TDH states that PWA resolve this conflict by choosing randomly between the two potential interpretations of the sentence. Thus, the TDH predicts that PWAs will perform at chance-level in the comprehension of sentences with non-canonical A-T word order (but cf. Caramazza, Capasso, Capitani, & Miceli, 2005; Caramazza, Capitani, Rey, & Berndt, 2001).

The abovementioned guessing interpretation is related to one important limitation of off-line sentence comprehension tasks. Off-line methods are static in the sense that they measure how participants interpret a sentence once its presentation has concluded, but they do not provide information related to the type of knowledge that listeners tap into to achieve a specific interpretation. The introduction of online methodologies has made an important contribution to aphasiology not only by addressing some of the limitations of off-line methods, but also by suggesting new interpretations of sentence comprehension data in PWA. First, we will briefly introduce the methodological framework.

<sup>14</sup> According to the Government and Binding model (GB; Chomsky, 1981), the displacement of a sentence constituent leaves behind a trace in its base-generated position. Thus, sentence comprehension requires keeping track of both the element in the derived position and the trace left in the base-generated position.

#### 4.1.2 Eye-tracking studies on sentence comprehension deficits

To examine real time sentence processing, several studies have used eye-tracking (ET) technology in the Visual World Paradigm (VWP). The VWP is based on the idea that the linguistic stimuli mediate visual attention shifts within a visual display due to referentially-driven processes (Cooper, 1974; Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995; see Tanenhaus, 2007 and Boland, 2004 for an overview). Participants' eye fixations on pictures in a scene are monitored as participants listen to sentences. Changes in the location and timing of fixations along the time string reveal how visual attention shifts in response to the continuous auditory stimuli. Some findings suggest that language mediated eye gaze tends to be unconscious, but not fully automatic (Mishra, Olivers, & Huettig, 2013). Current data suggest that the VWP provides a sensitive measure of word order processing in sentence comprehension.

Dickey, Choy, and Thompson (2007) introduced the VWP in aphasiology research in a study on *wh*-questions. They found that the comprehension of non-canonical object *wh*-questions was at chance-level in PWA. Interestingly, gaze data analysis did not converge with the traditional interpretation of chance-level accuracy (i.e. guessing) (TDH; Grodzinsky, 1986, 1995, 2000; see Draai & Grodzinsky, 2006ab for a later revision; see Burchert, Hanne, & Vasishth, 2013 for a review). Online sentence processing uncovered different processing routines in PWA as a function of off-line response accuracy. That is, PWA showed distinct fixation patterns in the correctly and incorrectly answered trials. In the correctly answered trials PWA showed the same gaze fixation pattern of NBD participants, indicating rapid and automatic thematic role assignment. In contrast, in the incorrectly answered trials, PWA showed a completely different gaze fixation pattern, characterized by a progressive proportion of fixations towards the foil picture. It constituted the first empirical evidence against the guessing interpretation of chance-level performance in PWA. Similar results have been reported in subsequent studies combining VWP with a sentence-picture matching task in German (Hanne et al., 2011) and Basque (Arantzeta et al., 2017, Chapter 2) PWA. Converging with Dickey et al. (2007), unsuccessful off-line performance in PWA was guided by fixation patterns that diverge from correctly interpreted sentences. Altogether, evidence from online data does not support the existence of a guessing pattern in sentence processing in PWA, but distinctive parsing routines that determine the interpretation of the sentence. In successful interpretations of the sentence, PWA show a processing pattern comparable to NBD, but it is still unclear what processes underlie sentence misinterpretation.

Online data analysis of incorrect answers might provide insights regarding the intermittent and hardly predictable (i.e., stochastic) failure shown by PWA in processing sentences with non-canonical word orders. However, this analysis is often challenging due to the small sample sizes and noisy data (see Caplan et al., 2007, for an alternative across-

task/measures approach). Dickey et al. (2007) suggested that PWA fail to comprehend sentences with non-canonical word order due to their inability to inhibit counteractive information; for example, the influence of agent-first heuristics. This refers to the tendency to assign the agent role to the first DP in the sentence (Bever, 1970; Bornkessel-Schlesewsky & Schlewsky, 2013). This heuristic achieves correct interpretation of sentences presented in the canonical word order, also in healthy speakers (Ferreira & Patson, 2007; van Herten, Chwilla, & Kolk, 2006; Townsend & Bever, 2001), but it fails to correctly assign thematic roles in derived word orders. In those cases, listeners must revise the initial parsing by applying more effortful analytical computations. Hanne et al. (2011) suggested that PWA have an early preference for interpreting sentences as canonical or non-canonical based on a “deterministic parsing”, followed by the inability to revise the initial parsing computation, even when they detect the need for reanalysis. Altogether, the failure of PWA to inhibit the antagonist interpretation of a semantically reversible sentence has been pointed out as a potential cause of the inconsistent sentence processing failure in PWA. Still, it is an open question whether the use of heuristics by PWA is a consciously learned and self-initiated procedure to compensate for their linguistic deficits, or whether it reflects an unconscious breakdown of parsing routines.

#### 4.1.3 Consciousness and compensatory strategies in PWA

Provided that PWA do not show anosognosia (i.e., unawareness of the aphasic condition) (see Kertesz, 2010; Vuilleumier, 2004 for an overview), PWA use self-initiated compensatory strategies to overcome their communicative impairments. These might be external (e.g., ask for adaptations to the interlocutor, use of electronic devices) or internal (e.g., self-cuing, verbal repetition, mental association) (e.g., Beeke, Wilkinson, & Maxim, 2009; Oelschlaeger & Damico, 1998; Tompkins, Scharp, & Marshall, 2006; see Simmons-Mackie & Damico, 1997).

In general, the use of heuristics reduces processing time and effort in parsing routines when compared to analytical processes (Shah & Oppenheimer, 2008). There is evidence suggesting that the PWA have limited resource availability to process the linguistic stimuli (Caplan et al., 2007; Kolk, 1993; Miyake, Carpenter, & Just, 1994). Thus, PWA may adopt a conscious shortcut to diminish the cognitive load involved in the parsing process. Among the languages studied in sentence processing in PWA, none has Theme-Agent order as its canonical structure. Indeed, this is a rare pattern shown by less than 4% of the languages worldwide (Dryer, 2005). The frequency of appearance of structures is a primary criterion to determine word order typology across languages; thus, structures with non-canonical word order tend to be less frequent than structures with canonical word order (Dryer, 2007). Hence, if we consider the above factors, reliance on agent-first strategy may be considered as the “best guess” under an arbitrary degree of success. It is unknown to what

extent the adoption of an agent-first strategy by PWA is based on a conscious decision that aids comprehension efficiency on an everyday basis. The focus lies in establishing a threshold between the conscious and unconscious processing of language.

Metacognitive tasks assess self-awareness and, hence, consciousness (Seth, Dienes, Cleeremans, Overgaard, & Pessoa, 2008; see Peña-Ayala & Cárdenas, 2015 for an overview). Metacognitive awareness refers to the subjective perception of one's own cognitive process. It can be measured by the subjective confidence ratings self-reported by the subject in a given task (see Norman & Price, 2015). Participants perform a primary task (e.g., picture matching) and subsequently are asked to rate their confidence in the validity of their decision. The degree of correspondence between the objective performance (i.e., accuracy) and the subjective confidence rating is used to assess the extent to which the primary task is mediated by conscious knowledge (Cheesman & Merikle, 1984, 1986; Dienes, Altmann, Kwan, & Goode, 1995; Overgaard, Timmermans, Sandberg, & Cleeremans, 2010). Conscious cognition is strongly associated with voluntary control. However, voluntary actions become automatic with practice (Shiffrin & Schneider, 1977) and consequently, the cognitive control over them decreases (e.g., Langer & Imber, 1979; Schneider, 2009). Hence, it is important to keep in mind that there is some degree of contribution of conscious and unconscious knowledge across most cognitive tasks.

Note that a judgment based on confidence rating is certainly a constituent of conscious awareness, but it may be the product of an unconscious inference. What is important is that once the judgment is built, it may augment self-control, and therefore the degree of personal regulation over processes that would otherwise influence behavior directly (Koriat, 2000). Dienes et al. (1995) suggested two criteria to use confidence rates to discern between conscious and unconscious knowledge; the zero-correlation criterion and the guessing-criterion. The former refers to the lack of relationship between confidence rating and objective accuracy, while the latter refers to the observed above-chance performance in the primary task while participants express themselves to be guessing. According to the authors, the fulfilment of these two criteria is a strong indicator of unconscious processing. This paradigm has been used mostly in perceptual discrimination and implicit learning tasks (Norman & Price, 2015; Overgaard, 2015, for an overview). In the current study it is applied for the first time in an aphasiology study.

Contrary to production experiments (Marshall & Tompkins, 1982; Oomen, Postma, & Kolk, 2001), studies of metacognitive awareness in comprehension deficits have been mainly on jargon aphasia (Marshall, Rappaport, & Garcia-Bunuel, 1985; Shuren et al., 1995; see Rubens & Garret, 1991, for an overview). To the best of our knowledge, the only study on comprehension error awareness in non-fluent PWA was conducted by Kennedy and Chiou (2008), who studied a group of Broca's and anomic PWA regarding metacognitive awareness on discourse-related questions. Kennedy and Chiou (2008)

reported that metacognitive awareness was mainly explained by the linguistic abilities of the PWA (discourse comprehension and design fluency repetition score), but also executive functions, such as switching and perseveration, contributed to a lesser extent (see also Stuss, 1991). In line with the perceptual loop theory (Levett, 1989, 2001), deficits in linguistic abilities relate to an impaired perceptual loop, and consequently a gradient of self-monitoring abilities.

We aimed to study the extent to which people with non-fluent aphasia have conscious knowledge of comprehension processing. Comprehension accuracy and self-reported confidence ratings were considered during a comprehension task involving canonical and non-canonical sentences.

#### 4.1.4 Research questions

Behavioral data were collected by using a sentence-picture matching task, followed by a confidence rating. Moreover, participants were monitored with an eye-tracker while performing the primary task (i.e., sentence-picture matching task) as an online measure of sentence processing. The latter aimed to further explore, in a post-hoc analysis, the interaction between self-awareness and real time sentence processing. This study poses three research questions:

- (1) Are PWA aware of their sentence comprehension errors?

The confidence rates reported to the incorrectly answered trials were studied to uncover the extent to which PWA performed a correct judgment of their (failed) comprehension accuracy. Correct judgments of incorrectly answered trials were considered as an indicator of error awareness in sentence comprehension. In contrast, incorrect judgments of incorrectly answered trials suggest that PWA were not aware of their comprehension errors.

- (2) Is sentence comprehension performance unconsciously mediated in PWA and NBD?

Following Dienes et al. (1995), this research question was answered based on the zero-correlation criterion and the guessing-criterion. In relation to the zero-correlation criterion, no correlation between objective comprehension performance and subjective confidence rating is an indicator of unconscious knowledge. The opposite pattern would be indicative of consciously mediated processing, and therefore voluntary control over thematic-role parsing. The latter finding would support that metacognitive tasks are valuable to assess the what extent to which PWA use agent-first heuristics as explicit



inference to disentangle thematic roles. Regarding the guessing-criterion, above chance performance in trials wherein participants believed that they were guessing indicate unconscious knowledge (i.e., implicit knowledge) of sentence processing.

- (3) To what extent do PWA answer by guessing on a task for comprehension of sentences in non-linear argument order?

Descriptive analysis of the data was used to add further evidence on the validity of the Trace Deletion Hypothesis (TDH; Grodzinsky, 1986, 1995, 2000; see Draai & Grodzinsky, 2006ab for a later revision). As described before, the TDH states that PWA answer by guessing when confronted with sentences presented in derived word orders. Based on the definition of guessing as “form(ing) or express(ing) an uncertain estimate or conclusion (about something), based on insufficient information” (Collins English Dictionary, online version), we may establish that it refers to a conscious act. Hence, the “guessing” pattern will be quantified through the self-ratings of PWA on correctly and incorrectly judged sentences, and contrasted with the TDH.

## 4.2 Methods

This study was approved by the Basque Clinical Research Ethics Committee (CEIC-E). All participants, or legal tutors, gave informed consent according to the declaration of Helsinki.

### 4.2.1 Participants

Fourteen individuals (mean age 66,07 years; SD= 10,38; range= 55-85; male/female= 11:3) with chronic non-fluent aphasia were tested in this study. Half of them were native Spanish monolingual speakers, whereas the others were L1Basque-L2Spanish bilingual speakers<sup>15</sup>. They were all pre-morbidly right-handed, as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971), and presented with aphasia due to cerebrovascular injury. Visual neglect was excluded using the Behavioural Inattention Test (BIT; Wilson, Cockburn, & Halligan, 1987). Fourteen individuals without any history of neurological or sensory impairments composed the NBD group. They were matched on age range and bilingualism with the clinical group (mean age 62,92 years; SD= 12,04; range= 44-82; male/female= 8:6). Demographic, linguistic and clinical information is provided in **Appendix B1**.

---

<sup>15</sup> All bilingual participants had acquired Spanish at an early age (< 5 years) and were literate only in Spanish, their language of instruction at school.

PWA were screened in Spanish for word and sentence comprehension abilities using the extended version of the Boston Aphasia Test (BDAE; Goodglass, Kaplan, & Barresi, 2005; Adapted to Spanish by García-Albea, 2005). The subsection of conversation and language exposition, comprehension of words, commands, complex ideational materials and syntactic processing subsections were used. Furthermore, PWA were also assessed for working memory using the digit-span task of the Wechsler Adult Intelligence Scale (WAIS-III; Wechsler, 1997). Scores from the subtests of the BDAE and the WAIS-III are presented in **Appendix B3**.

To be included in the present study PWA showed relatively preserved lexical comprehension (> 80%) and impaired sentence processing (< 66.66%) as assessed by the “Touch A with B” subtest. This subtest distinguishes a reversible relationship between sentence constituents presented in a variety of word orders.

#### 4.2.2 Design and Materials

The materials used in this study were adapted to standard European Spanish from the experiment presented in Chapter 2. They consisted of single sentences provided auditorily in combination with the presentation of two pictures on the screen. Each of the pictures depicted two people taking part in the performance of the same action, but with reverse Agent-Theme thematic roles (see **Figure 3.1** in Chapter 3). There were 126 trials consisting of 120 experimental items and 6 practice items.

##### 4.2.2.1 Picture-matching task

###### *Linguistic stimuli:*

The same twenty-two transitive verbs used in the original study (Arantzeta et al., 2017), as well as two animate singular DPs assigned to each verb were selected to create declarative sentences in the following structures (a) active; (b) passive; (c) subject relative; (d) object relative; (e) subject cleft; (f) object cleft. Subsequently, sentence conditions were clustered as Agent-Theme (A-T) and Theme-Agent (T-A) for data analysis. The Agent-Theme assignment in the DPs was randomized and balanced within the six conditions. That is, each DP was the Agent of the sentence in three out of six conditions. Regions of interest (ROIs) of the experimental stimuli were individually measured using the Computerized Language Analysis software (CLAN; MacWhinney, 2000). Unlike the original version in Basque (Arantzeta et al., 2017), the constituents of the sentences did not have the same length across experimental items. Nonetheless, t-test comparisons of ROI durations between paired conditions (i.e., active vs. passive, subject vs. object relative, subject vs. object cleft) showed no difference (see **Table 4.1**). This was essential for subsequent comparison of temporal data (i.e. gaze data) across paired conditions.

**Table 4.1.** Regions of Interest (ROI), duration (mean and SD) and comparison of length across paired conditions.

Paired conditions	ROI (mean duration and SD)					
	ROI 1 Argument 1 999 ms (46)		ROI 2 Verb 931 ms (12)		ROI 3 Argument 2 852 ms (30)	
	t	p	t	p	t	p
Active versus Passive	0.945	0.350	-0.174	0.863	-1.376	0.177
Subj. versus Obj. Relative	-0.221	0.826	-1.877	0.068	1.268	0.213
Subj. versus Obj. Cleft	1.074	0.290	-0.397	0.693	0.366	0.716
A-T versus T-A	-0.554	0.581	-0.027	0.177	0.077	0.939

ROI= Region Of Interest; A-T= Agent-Theme; T-A= Theme-Agent

In the linguistic stimuli the active sentences (1) were constructed with perfect present tense to keep the length of the verb as equal as possible with the counterpart passive sentences (2). In the relative clauses the antecedent was always introduced by the verbal phrase “I see” and the relative pronoun “que” functioning as subject (3) or object (4). In the cleft sentence constructions, the contrastive element became the complement of the copular verb “ser”, and the relativizer pronoun “que” introduced the rest of the sentence (5-6). In both object-cleft and object-relative constructions, the relative pronoun was preceded by the preposition “a”, as well as the direct object in the cleft constructions.

- (1) La mujer ha peinado a la niña  
 det woman aux.has comb-PTCP prep det girl  
*The woman has combed the girl.*
- (2) La niña ha sido peinada por la mujer  
 det girl aux.has be-PTCP comb-PTCP prep det woman  
*The girl has been combed by the woman.*
- (3) Veo a la mujer que peina a la niña  
 see prep det woman pron-rel comb prep det girl  
*I see the woman who combs the girl.*
- (4) Veo a la niña a la que peina la mujer  
 see prep det girl prep det pron-rel comb det woman  
*I see the girl who the woman combs.*
- (5) Es la mujer la que peina a la niña  
 be det woman det rel-pron comb prep det girl  
*It is the woman who combs the girl.*

- (6) Es a la niña a la que peina la mujer  
 be prep det girl prep det rel-pron comb det woman  
*It is the girl who the woman combs.*

A female native speaker of standard peninsular Spanish recorded the linguistic stimuli in a soundproof booth (IAC) using a digital microphone (audio-technica AT4022a). All sentences were recorded with a constant prosodic contour to avoid biased thematic role interpretations based on intonation (Ferreira, Anes, & Horine, 1996; Weber, Grice, & Crocker, 2006).

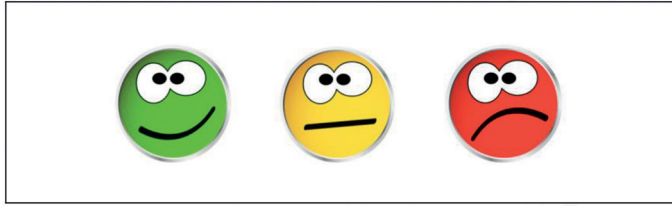
#### *Visual stimuli*

The 44 black-and-white line drawings used by Arantzeta et al. (2017) were adapted to the current study. The drawings constituted 22 pairs, which were presented together divided by a black vertical line in the middle of the screen. The two pictures within each pair showed the same action with an Agent-Theme reversal. Each drawing measured approximately 15x15 cm. See Arantzeta et al. (2017) for detailed information on the visual stimuli and their corresponding normalization.

The presentation of the experimental items in the visual display (i.e. right vs. left) was pseudo-randomized. No more than two target stimuli could occur in a row on the same side of the screen. The direction in which the action was performed was also balanced across the stimuli to avoid rightward-inclined scanning (Scheepers & Crocker, 2004).

#### *4.2.2.1 Confidence rating*

A three-interval confidence rating was used and visually represented by coloured emoticons. From right to left, a green smiley face, an amber neutral face and a red sad face were presented following a horizontal axis (see **Figure 4.1**). Each emoticon measured approximately 10x10 cm. The first (i.e., green smiley emoticon) and last (red sad emoticon) implied confidence of response to some degree, corresponding to “sure I answered correctly” and “sure I answered incorrectly”, respectively. The middle response alternative (i.e., amber neutral) corresponded to having no knowledge to report the accuracy judgment, and it was defined as “I don’t know/guess”.



**Figure 4.1.** Response grill used to self-report confidence rate on the accuracy of the sentence-picture matching task. The emoticons represented the next confidence rates: green smiley face= “sure I answered correctly”; amber neutral face= “I don’t know/guess”; red sad face= “sure I answered incorrectly”.

### 4.2.3 Procedure

The experiment was programmed using E-Prime 2.0.10 with extensions for Tobii 2.0.2.41 (ClearView; Psychology Software Tools, Pittsburgh, PA). The visual stimulus was presented on a 23” wide screen LED monitor having 1280\*720 resolution. The auditory stimuli were delivered binaurally through headphones. Participants were seated 60–70 cm from the screen, with a visual angle under 15° (max. Allow 35°), while their eye movements were captured using a Tobii 120 Desktop Eye Tracker (sampling rate 120 Hz, accuracy 0.5 degrees).

The presentation of the experimental stimuli was divided into four blocks of 30 items. Two blocks were administered in each of the two experimental sessions, after the presentation of six practice items. No more than two consecutive items corresponded to the same linguistic condition.

Prior to the start of the presentation of each block of stimuli, a 5-point calibration sequence was used to calibrate the eye-tracker to each subject. Subsequently, written instructions for the task appeared on the screen. The same instructions were also read aloud and verbally explained to all participants.

The experiment consisted of two consecutive tasks. In the primary task, the participants were asked to perform a picture-matching task, while in the secondary task they were asked to report the perceived correctness of their previous answer using a confidence rate. Trial-by-trial subjective measures were applied, since they have been proven to be more sensitive than post-task reports (Ziori & Dienes, 2006).

At the beginning of each trial, a fixation smiley face was centred in the screen. Participants had to fixate on the image for 250 ms before it was replaced by the display of a pair of pictures corresponding to the experimental stimuli. After 1000 ms of its previsualization, the auditory stimulus was presented. Participants were asked to select the picture that best depicted the auditorily presented sentence by pressing specific keys on the keyboard with the left hand<sup>16</sup>. Trials with no response within 8000 ms from the

<sup>16</sup> As an exception, a participant with crossed aphasia (A2) answered with the right-hand.

offset of the sentence were counted as having no response. When the decision was made, the second task was introduced. Participants were asked to make a confidence rating, as an assessment of the accuracy of their response on the previous task (i.e. picture-matching task). To this end, the three colour emoticons appeared on the screen and participants had to select the one that best represented their judgment. They were allowed to make their choice by pressing specific buttons on the keyboard or pointing directly on the screen. Confidence ratings provided by pointing directly on the screen were entered by the researcher using a second keyboard. The participants had a maximum of 12000 ms to answer. After this time, the confidence rating was registered as not being answered and the next trial was started. Trials answered before the onset of the auditory presentation (i.e., previsualization face) were excluded from subsequent data analysis, corresponding to the 0.25% of the total data, since they may correspond to accidental button press.

In both the first and second tasks, trials with no responses were excluded from further analysis, corresponding to 1.28 % and 0.58% of the total data, respectively. Fixations with durations less than 90ms were removed from the analysis to exclude ocular artefacts (e.g., blinks and saccades). We introduced a switch of 200 ms to correct for the time required for planning and executing an eye-movement, and thereby, to time align the gaze fixation data with the linguistic stimuli (Matin, Shao, & Boff, 1993).

### 4.3 Data analysis

In addition to descriptive statistics, we used regression analyses to examine the relationship between experimental variables and predictors. Different types of regression analyses were used depending on the type of dataset and the related research questions.

Comprehension accuracy data and post-hoc gaze data analysis were conducted using Generalized Linear Mixed-effects Models (GLMM) with logit function, and Linear Mixed-effects Models (LMM), respectively. As a statistical technique, (G)LMM combines fixed-effects terms (i.e., regression coefficients) and random-effects terms. The former defines the expected values of the observations, while the latter introduces the variance and covariance of the observations of the subjects and linguistic stimuli (see Bates et al., 2005). Mixed models are resistant to the impact of outlying observations and missing data, common characteristics of small sample sizes and longitudinal data, respectively (see Verbeke & Molenberghs, 2000; Diggle, Heagerty, Liang, & Zeger, 2002). For a detailed description of GLMM and LMM techniques the reader is directed to McCulloch and Searle (2000).

Model building was conducted by progressively introducing random-effects, fixed effects and corresponding interactions. Akaike's Information Criterion (AIC; Akaike,

1974) was used to measure the goodness of fit and compare models with each other. The model with the lowest AIC value was considered to fit significantly better only when the difference between the two AICs was equal to or higher than two. The numerical predictors Age and Trial number were centred. Least-square means (LSMeans) were used for comparing LS-mean differences on the basis of the mixed model.

Specific procedures were followed to test the zero-correlation and guessing criterions. Zero-correlation criterion was analyzed with simple Logistic Regression Models (LRM) fitted with a logit link function, in addition to the estimation of the area under the receiver operating characteristic (ROC) curve. The area under the curve (AUC) is a measure of the discrimination of the predictors, which ranges from 0.5 indicating random discriminative ability, to 1.0, indicating perfect discrimination. Thus, the area under the curve is the probability that participants correctly rate their accuracy. The ROC curve analysis has been widely used for diagnostic purposes (Zweig & Campbell, 1993; see Streiner & Cairney, 2007 for a review). The nonparametric approach of DeLong (DeLong, DeLong, & Clarke-Pearson, 1988) and bootstrapping methods were used to compare the AUCs between groups. In addition, the analysis was complemented with Spearman's correlation coefficient to measure the strength of the relationship between the objective accuracy and the subjective confidence rate provided by the participants.

The guessing criterion was examined using Linear Regression Models for binary data (i.e., GLMM), by comparing the linear predictor with zero. LSMeans (also called model-adjusted means) and 95% CIs were used for this purpose.

Tukey correction was used for multiple comparisons, and pairwise differences at  $p < .05$  were considered significant, unless otherwise indicated. The statistical software R was used for this analysis (R Core Team, v.3.2.3.) with the next packages; lme4 (Bates et al., 2015), rms (Harrell, 2016), and pROC (Robin et al., 2011).

## 4.4 Results

### 4.4.1 Comprehension accuracy

Sentence conditions were clustered as Agent-Theme (A-T) and Theme-Agent (T-A). The former included active, subject cleft and subject relative conditions, while the latter included the counter pairs: passives, object clefts and object relatives.

PWA currently comprehended 79% (Standard Error,  $SE = 1.42$ ) of the sentences in A-T order and 65% ( $SE = 1.67$ ) of the sentences presented in T-A order. The NBD group performed at ceiling level, obtaining 96% ( $SE = 0.69$ ) accuracy in A-T and 96% ( $SE = 0.64$ ) accuracy in T-A argument orders. Detailed statistics on sentence comprehension scores

by argument orders are presented in **Table 4.2**. (See **Appendix B5** for individual participants' scores).

**Table 4.2.** Comprehension accuracy (%) and standard error (SE) as a function of group and sentence condition.

Group	Condition					
	Agent-Theme			Theme-Agent		
	Active	Subj. Cleft	Subj. Rel.	Passive	Obj. Cleft	Obj. Rel
PWA	81.88%	79.92%	74.81%	74.34%	60.44%	61.62%
	(2.32)	(2.44)	(2.60)	(2.66)	(2.99)	(2.95)
NBD	98.17%	96.05%	93.18%	97.45%	96.78%	95.00%
	(0.80)	(1.16)	(1.51)	(0.95)	(1.05)	(1.30)

PWA= people with aphasia; NBD= non-brain-damaged

The final GLMM obtained for the accuracy data consisted of a three-way interaction for group, argument order and trial number, and stimuli and subject variables as random-effects.

The PWA performed significantly poorer than the NBD across both A-T ( $\beta = -2.029$ ;  $SE = 0.324$ ;  $p < .0001$ ) and T-A ( $\beta = -2.703$ ;  $SE = 0.305$ ;  $p < .0001$ ) argument orders. The PWA comprehended sentences presented in A-T better than T-A argument order ( $\beta = 0.754$ ;  $SE = 0.120$ ;  $p < .0001$ ), while the people with NBD showed no effect regarding word order ( $\beta = 0.080$ ;  $SE = 0.277$ ;  $p = 0.7728$ ). In addition, there was an interaction between group, argument order and trial number (i.e., position of the presentation of a given trial in relation to the others). The NBD participants became more accurate in the sentence picture matching task across the course of the experiment, but this improvement was restricted to sentences presented in A-T argument order (A-T;  $\beta = -2.43$ ;  $SE = 0.704$ ;  $p = 0.0006$ . T-A;  $\beta = -0.736$ ;  $SE = 0.701$ ;  $p = 0.2942$ ). The PWA did not show any effect of trial number across any of the argument orders (A-T;  $\beta = -0.414$ ;  $SE = 0.343$ ;  $p = 0.2278$ . T-A;  $\beta = -0.327$ ;  $SE = 0.330$ ;  $p = 0.3229$ ).

#### 4.4.2 Confidence ratings

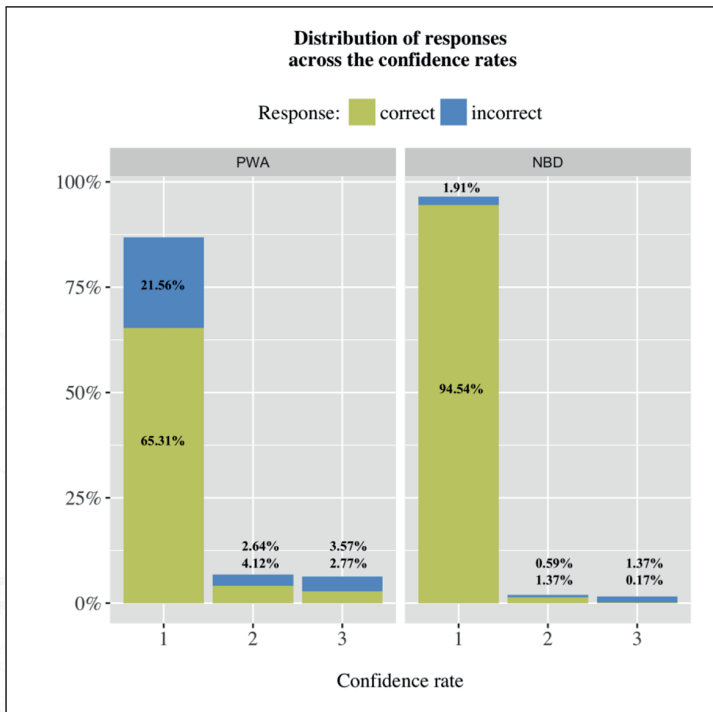
Participants expressed their confidence about the correctness of their answer in the sentence-picture matching task by choosing within three options. The distribution of the confidence rates is presented in **Table 4.3**, separately for correctly and incorrectly comprehended trials. See **Figure 4.2** for an overall distribution of responses in each group.



**Table 4.3.** Distribution of confidence ratings on comprehension accuracy in the primary task (i.e., sentence-picture matching task) as a function of response accuracy and group.

Response accuracy:	PWA			NBD		
	1	2	3	1	2	3
<b>Correct answers</b>	90.44%	5.71%	3.83%	98.37%	1.43%	0.18%
	(n= 1060)	(n= 67)	(n= 45)	(n= 1576)	(n= 23)	(n= 3)
<b>Incorrect answers</b>	77.60%	9.53%	12.86%	49.23%	15.38%	35.38%
	(n= 350)	(n= 43)	(n= 58)	(n= 32)	(n= 10)	(n= 23)

PWA= people with aphasia; NBD= non-brain-damaged; 1= "sure I answered correctly"; 2= "I don't know/guess"; 3= "sure I answered incorrectly".



**Figure 4.2.** Distribution (%) of the self-reported judgment of the sentence-picture matching task as a function of response accuracy in both NBD and PWA groups.

Confidence rate 1= "sure I answered correctly", 2= "I do not know, guessing", 3= "sure I answered incorrectly". NBD= non-brain-damaged; PWA= people with aphasia.

#### 4.4.3 Zero-correlation criterion

Data were arrayed into correct/incorrect judgments (i.e., binary data) for clarification, as provided in **Table 4.4**. Correct judgments refer to the instances in which participants reported as correct the trials answered correctly and reported as incorrect the trials answered incorrectly. Thus, regardless of the accuracy of their response in the primary task (i.e., sentence-picture matching task), participants made a correct judgment of their performance. Incorrect judgments refer to the cases in which participants' subjective self-report (i.e., confidence rating) did not match with the correctness of their answer in the sentence-picture matching task.

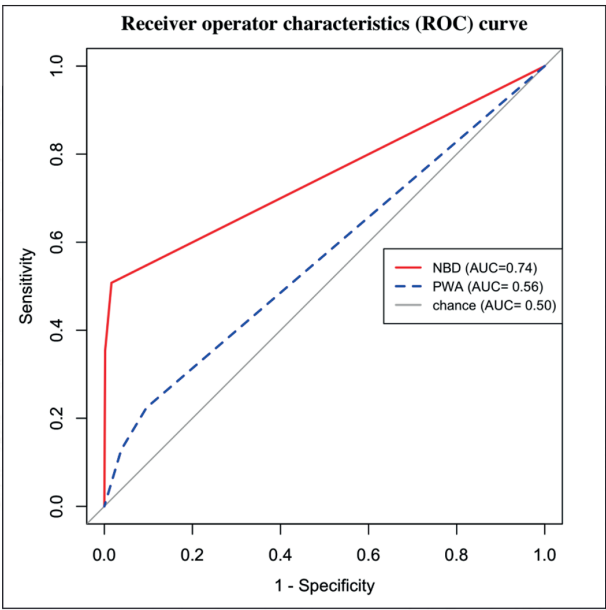
**Table 4.4.** Proportion of correct and incorrect judgments over the trials in which participants expressed themselves as being sure about their answer on the confidence rate (i.e., "sure I answered correctly" and "sure I answered incorrectly"). Calculations across groups and, correctly and incorrectly answered trials.

Response accuracy:	PWA		NBD	
	C. judgment	I. judgment	C. judgment	I. judgment
<b>Overall</b>	73.89%	26.10%	97.85%	2.14%
<b>Correct answer</b>	95.92%	4.07%	99.81%	0.18%
	(n= 1060)	(n= 45)	(n= 1576)	(n= 3)
<b>Incorrect answer</b>	14.21%	85.78%	41.81%	58.18%
	(n= 58)	(n= 350)	(n= 23)	(n= 32)

PWA= people with aphasia; NBD= non-brain-damaged; C. judgment= correct judgment; I. judgment= incorrect judgment.

Simple Logistic Regression Model (LRM) analysis was conducted to explore the relationship between the outcome of sentence comprehension accuracy and the confidence rating provided by the participants. Separate logit models were fitted for each group of participants. ROC analysis considered together the sensitivity (percentage of correctly identified incorrect answers) and the specificity (percentage of incorrectly identified correct answers) across a range of values for the ability to predict the sentence accuracy outcome. The results are illustrated in **Figure 4.3**. The area under the curve (AUC) for the confidence rating that perfectly predicts comprehension accuracy would be 1. In our sample data, the confidence rating had an AUC of 0.56 (95% CI: 0.54–0.58) in the PWA group, just better than random (i.e., 0.50), while the group with NBD had an AUC of 0.74 (95% CI: 0.68–0.81). A comparison of AUCs conducted by DeLong's method uncovered significant differences between the groups ( $p < .0001$ ), which was confirmed by bootstrapping analysis (100,000 samples;  $p < .0001$ ). Spearman's correlation coefficient

between comprehension accuracy and confidence rating was 0.1747 in the PWA group and 0.5175 in the NBD group. The former corresponded to a very weak effect size, while the latter was moderate.



**Figure 4.3.** Predicted probability of the confidence rating on the comprehension accuracy, illustrated by the sensitivity (i.e., true positive rate) against 1-Specificity (i.e., false positive rate).

NBD= non-brain-damaged; PWA= people with aphasia. The area under the ROC curve (AUC) is a measure of how well confidence ratings can distinguish between correctly and incorrectly answered trials.

#### 4.4.4 Guessing-criterion

The proportion and distribution of the instances in which participants reported having guessed are presented in the contingency table in **Table 4.5**. Participants in this study seldom reported they were guessing. Nonetheless, when both PWA and people with NBD reported answering randomly, without enough information, the response accuracy was above the chance baseline (i.e., 50%).

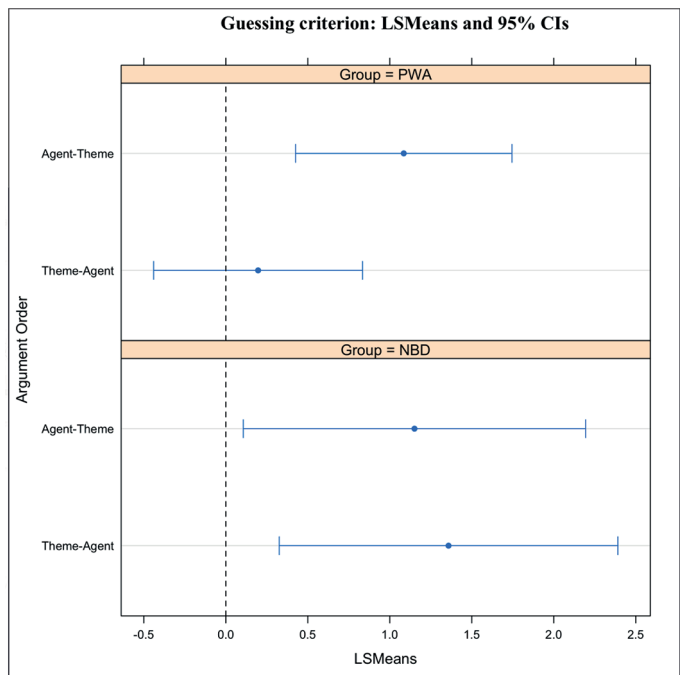
**Table 4.5.** Distribution of the instances in which participants declared themselves to be guessing, across correctly and incorrectly answered trials in both PWA and NBD groups.

Response accuracy:	PWA		NBD	
	Num. trials (n=110)	%	Num. trials (n=33)	%
Correct answer	67	60.90%	23	69.69%
<b>Incorrect answer</b>	43	39.09%	10	30.30%

PWA= people with aphasia; NBD= non-brain-damaged.

The empirical model consisted of a two-way interaction between the group and confidence rating and a two-way interaction between group and argument order, in addition to random-effects for subject and stimulus. A nested random-effect was included to account for the effect of sentence condition and argument order in each stimulus and subject, respectively.

The guessing criterion was studied by comparing the linear predictor with zero. The regression intercept was significantly positive in PWA (LSMean= 0.640; SE= 0.305; 95% CI= 0.04–1.23;  $t(109)= 2.098$ ;  $p= 0.0382$ ) and those with NBD (LSMean= 1.254; SE= 0.490; 95% CI= 0.29–2.21;  $t(32)= 2.557$ ;  $p= 0.0155$ ), indicating above-chance accuracy. The same analysis was conducted separately as a function of argument order, as illustrated in **Figure 4.4**. In the PWA group, the lower limit of the 95% CI was significantly positive in sentences presented in A-T argument order (LSMean= 1.085; SE= 0.336; 95% CI= 0.42–1.74;  $t(44)= 2.339$ ;  $p= 0.0239$ ), but not in sentences presented in T-A argument order (LSMean= 0.196; SE= 0.325; 95% CI= -0.44–0.833;  $t(64)= 0.605$ ;  $p= 0.5469$ ). In the group of people with NBD, the intercept was significantly positive across sentence presented in A-T argument order (LSMean= 1.150; SE= 0.532; 95% CI= 0.10–2.19;  $t(14)= 2.16$ ;  $p= 0.0486$ ), as well as in T-A argument order (LSMean= 1.358; SE= 0.526; 95% CI= 0.32–2.39;  $t(17)= 2.579$ ;  $p= 0.0195$ ).



**Figure 4.4.** Illustration of the least-square means (LSMeans) and 95% confidence intervals across groups and argument orders in trials where participants declared themselves to be answering by guessing.

PWA= people with aphasia; NBD= non-brain-damaged.

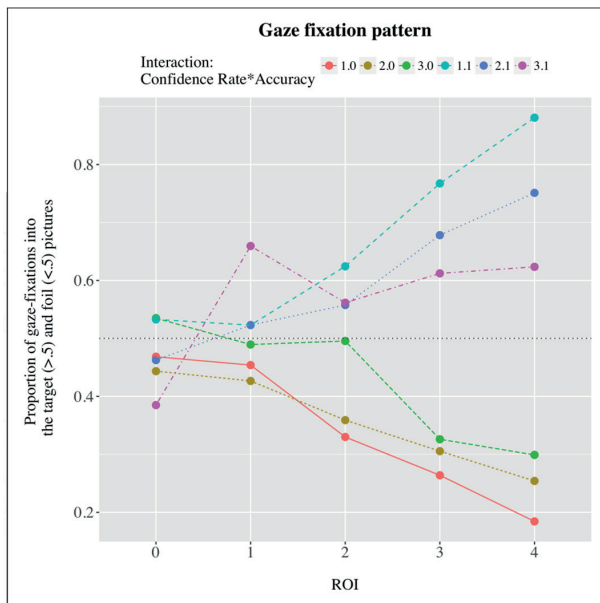
#### 4.4.5 Gaze data analysis

The difference in the proportion of fixations between the target and foil pictures was computed across different temporal Regions Of Interest (ROIs). ROIs 1, 2 and 3 corresponded to the presentation of the first, second and third constituents of the sentence. In addition, a post-offset ROI 4 with a duration of 1120 ms was included in the data analysis to account for gaze fixation following the end of the auditory presentation of the stimuli. As detailed in **Table 4.1**, the duration of the ROIs did not differ across pairs of sentence conditions, nor across A-T and T-A argument orders. This property of the experimental design provided the possibility to merge the gaze data across sentence conditions to gain statistical power to analyse the effect of response accuracy and confidence rate in real time sentence processing. Missing gaze data were imputed in the trials answered before the offset of ROI4 (i.e. mean RT < 3902 ms) based on the accuracy of the response.

Post-hoc analysis of the gaze data was conducted exclusively on the PWA, since the accuracy and confidence rating data were almost completely confounded in the NBD group, as shown in **Table 4.3**. The empirical model consisted of a three-way interaction between the response accuracy, ROI and confidence rating, in addition to random-effects for subject and stimulus. A Bonferroni procedure was used to control the error rate in testing multiple comparisons. The adjustment was applied, dividing the  $p$  level by the number of levels of the confidence rating (i.e., 3). Thus, an  $\alpha$  level of 0.016 was used.

The interaction between the independent variables is illustrated in **Figure 4.5**. Note that in all of the sentences, the target picture could in principle be determined on the basis of the first constituent (i.e., ROI 1) in subject/object cleft structures, and on the basis of the second constituent (i.e., ROI 2) in the rest of the sentence conditions. In line with previous studies, gaze fixation patterns along the visual display differed depending on the correctness of the answer from ROI 2 (see Arantzeta et al., 2017; Hanne et al., 2011). In correctly comprehended sentences, PWA showed an incremental proportion of fixations towards the target picture through the presentation of the sentence, while in the incorrectly answered trials PWA showed the opposite pattern. In addition to this, in the current study the confidence rating uncovered a distinctive gaze fixation pattern as a function of response accuracy.

The following results are based on gaze fixation differences along the visual display during the presentation of the linguistic stimuli as a whole. A detailed comparison of gaze data along specific ROIs is available in **Table 4.6**. In the correctly answered trials, the gaze fixation pattern for sentences that PWA reported to have answered correctly (i.e., confidence rating (CR)= 1) and sentences that PWA reported to have answered incorrectly (i.e., CR= 3) were distinctive ( $\beta = 0.084$ ;  $SE = 0.030$ ;  $p = 0.0060$ ). Trials rated with CR= 1 (i.e., “sure I answered correctly”) also corresponded to a different fixation pattern compared to trials in which PWA reported to answer by guessing (i.e., CR= 2) ( $\beta = 0.071$ ;  $SE = 0.024$ ;  $p = 0.0033$ ). Nonetheless, the fixation pattern of the latter was indistinguishable from incorrectly judged trials (i.e., CR= 3) ( $\beta = 0.013$ ;  $SE = 0.038$ ;  $p = 0.7305$ ). In the incorrectly answered trials, gaze fixation pattern was significantly different between trials rated as correctly understood (i.e., CR= 1) and trials rated as incorrectly understood (i.e., CR= 3) by PWA ( $\beta = -0.93$ ;  $SE = 0.028$ ;  $p = 0.0008$ ). There was no difference in the fixation patterns of the trials where PWA declared themselves to have guessed (i.e., CR= 2) versus trials answered with certainty (perceived as correct;  $\beta = -0.027$ ;  $SE = 0.031$ ;  $p = 0.3847$ ; perceived as incorrect;  $\beta = -0.066$ ;  $SE = 0.039$ ;  $p = 0.0959$ ).



**Figure 4.5.** Fixation pattern of PWA across the visual display along the presentation of the auditory stimuli. The gaze data is presented as a function of confidence rate (1,2,3) and accuracy (0,1).

Confidence rate, 1= “sure I answered correctly”, 2= “I do not know, guessing”, 3= “sure I answered incorrectly”. Accuracy, 0= incorrectly answered trials; 1= correctly answered trials.

Gaze fixation patterns for correctly and incorrectly answered trials were more-or-less mirror images. The data suggest that regardless of whether PWA answered correctly or not, when they were confident about the correctness of their answer (i.e., “sure I answered correctly”), they usually looked towards the target or foil picture from ROI 2. As soon as they were provided with an information cue that might guide thematic role parsing, they chose an interpretation and stayed with it. There is no indication that PWA doubt their choice when they get to the trace position.

**Table 4.6.** Comparison of gaze data per each ROI across the confidence rates provided by the PWA group.

	$\beta$	SE	z-ratio	p		$\beta$	SE	z-ratio	p
<b>Correct answers:</b>									
<b>C.R.</b>	<b>ROI<sub>1</sub></b>				<b>ROI<sub>2</sub></b>				
<b>1-2</b>	-0.000	0.048	-0.013	0.9897	0.066	0.048	1.366	0.1719	
<b>1-3</b>	-0.137	0.061	-2.238	0.0252	0.062	0.061	1.020	0.3074	
<b>2-3</b>	-0.136	0.076	-1.792	0.0732	-0.003	0.076	-0.051	0.9596	
<b>C.R.</b>	<b>ROI<sub>3</sub></b>				<b>ROI<sub>4</sub></b>				
<b>1-2</b>	0.088	0.048	1.841	0.0656	0.1293	0.047	2.696	<b>0.0070</b>	
<b>1-3</b>	0.154	0.061	2.520	<b>0.0117</b>	0.2567	0.060	4.236	<.0001	
<b>2-3</b>	0.065	0.076	0.861	0.3895	0.1274	0.075	1.688	0.0914	
<b>Incorrect answers:</b>									
<b>C.R.</b>	<b>ROI<sub>1</sub></b>				<b>ROI<sub>2</sub></b>				
<b>1-2</b>	0.028	0.064	0.437	0.6618	-0.028	0.063	-0.446	0.6550	
<b>1-3</b>	-0.034	0.055	-0.626	0.5310	-0.165	0.056	-2.940	<b>0.0033</b>	
<b>2-3</b>	-0.062	0.079	-0.787	0.4307	-0.136	0.079	-1.716	0.0862	
<b>C.R.</b>	<b>ROI<sub>3</sub></b>				<b>ROI<sub>4</sub></b>				
<b>1-2</b>	-0.040	0.062	-0.649	0.5161	-0.069	0.062	-1.103	0.2700	
<b>1-3</b>	-0.061	0.056	-1.099	0.2718	-0.114	0.056	-2.017	0.0437	
<b>2-3</b>	-0.020	0.079	-0.264	0.7918	-0.044	0.079	-0.564	0.5726	

C.R.= Confidence rate; 1= "sure I answered correctly"; 2= "I don't know/guess"; 3= "sure I answered incorrectly". ROI= Region Of Interest; ROI 1= first constituent of the sentence; ROI 2= second constituent of the sentence; ROI 3= Third constituent of the sentence; ROI 4= Post-offset region; Significance level  $p < .016$

We have analysed a group of PWA and NBD performing a sentence-picture matching task followed by a confidence rating. The accuracy data show that PWA were more likely to be correct in understanding sentences presented in A-T than in T-A argument order. NBD showed ceiling accuracy across all argument orders. The analysis of the response accuracy and confidence rating has shown that PWA tend to judge as correct both correctly and incorrectly comprehended trials; they were sure that they were correct on 90% of correct responses and 78% of incorrect responses. That is, PWA have shown little awareness of their comprehension errors. Confidence rating was a very weak predictor of comprehension accuracy in PWA; it was worse than in the NBD group. Both PWA



and NBD participants reported guessing in a small number of cases (6.8% and 2.0% of trials, respectively). In terms of the guessing criterion, when participants reported that they were guessing, the comprehension accuracy was above chance in both PWA and NBD groups. This pattern was not consistent in sentences presented in T-A order to the PWA group. Post-hoc analysis showed that both response accuracy and confidence rating were significant predictors of the gaze fixation data. Participants showed distinct fixation patterns depending on the confidence rating subsequently given to the trial.

## 4.5 Discussion

The questions addressed in this study are: (a) Are PWA aware of their sentence comprehension errors?; (b) Is sentence comprehension performance unconsciously mediated in PWA and NBD?; and (c) Do PWA guess in the comprehension of sentences with derived word orders?. In addition, we conducted a gaze data analysis to explore the relationship between real time sentence processing and the metacognitive awareness reported by PWA. This section will initially discuss sentence comprehension accuracy data to contextualize the subsequent discussion.

As expected, PWA taking part in this study performed worse than NBD in the sentence comprehension task. PWA comprehended sentences with a linear A-T order, better than the derived T-A order. The NBD group comprehended sentences in both A-T and T-A argument orders equally well. Altogether, the present findings converge with those of Chapter 2, whose linguistic and visual materials were adapted for the present study, as well as with previous studies involving Spanish speakers with aphasia (Juncós-Rabadán et al., 2009), and other languages (Swahili, Abuom, Shah, & Bastiaanse, 2013; German, Burchert et al., 2003; Turkish, Duman, Altinok, Özgirgin, & Bastiaanse, 2011; Italian, Garraffa & Grillo, 2008; English, Meyer et al., 2012; Schwartz et al., 1980; cf. in Indonesian, Jap, Martínez-Ferreiro, & Bastiaanse, 2016).

Metacognitive awareness of sentence comprehension accuracy differed significantly between correct and incorrect responses. In the former, both NBD and PWA usually made a correct judgment of their answer. That is, they perceived as correct the trials answered correctly. PWA frequently rated incorrect trials as correct. They failed to detect the error (i.e., false negative detection). Hence, PWA were generally not aware of their sentence comprehension difficulties. NBD made few comprehension errors, but similarly only a reduced part of these errors were correctly rated (i.e., 35% of the incorrectly answered trials). This indicates that not even NBD participants are always aware of comprehension failure, suggesting that it is not only the neurological injury in PWA that is impacting

awareness of error. This is a potentially interesting experimental question, but it is beyond the immediate scope of this study.

#### 4.5.1 Zero-correlation criterion

Following Dienes et al. (1995), the degree of correspondence between the accuracy and the confidence rating was used to assess the conscious versus unconscious threshold in sentence comprehension. The consistency between the measures varied across the two groups, suggesting that the extent to which sentence comprehension is mediated by conscious knowledge varies as a consequence of neurological injury. In the PWA group, the strength of the subjective perception (i.e., confidence ratings) in predicting the comprehension accuracy was slightly above chance, while in the NBD it was significantly larger, although not perfect. This result was confirmed by a very low correlation between the accuracy and confidence rating measures in the PWA group, and a moderate correlation in the NBD group.

The results suggest that PWA and NBD do not share the same consciousness threshold in sentence processing. In the PWA group, sentence comprehension is mainly mediated by unconscious knowledge, suggesting little voluntary control (i.e., dominance of automatic control). Consequently, PWA do not consciously perceive failure in the parsing routine. In contrast, in the NBD group, a moderate relationship between accuracy and confidence rating suggests that conscious and unconscious knowledge equally interplay in sentence processing. Thus, voluntary and automatic control appear to function together in thematic-role mapping in NBD, who appear to self-monitor, contrary to PWA. Note that the lack of control shown in the PWA group cannot be explained by practice (Shiffrin & Schneider, 1977; Langer & Imber, 1979), since agent-first heuristics have also been proven to be deployed in healthy speakers (Bornkessel-Schlesewsky & Schlesewsky, 2013; Ferreira & Patson, 2007; van Herten et al., 2006). Still, NBD show a moderate-good perception of the parsing mechanism, and reanalyze the sentence based on analytical routines (i.e., by processing morphological cues) when the use of heuristics fails to correctly interpret the sentence. It is not clear whether PWA do not perceive the need to reanalyze or whether the impairment lies in the reanalysis process *per se*. We first need to introduce the guessing-criterion to address this question.

#### 4.5.2 Guessing-criterion

The analysis of the instances in which participants said they were guessing uncovered two main things. First, both PWA and NBD rarely reported that their answer in the comprehension task was based on uncertainty. This finding does not support the TDH (TDH; Grodzinsky, 1986, 1995, 2000; see Draí & Grodzinsky, 2006ab for a later revision), which states that PWA choose randomly when thematic roles need to be assigned in

non-linear order. In contrast to the predictions of the TDH, PWA nearly always tended to provide confidence rates (correct or incorrectly) based on certainty (i.e., “sure I answered right/wrong”) – they did not report they were guessing. Second, according to the guessing-criterion (Dienes et al., 1995), the two groups performed above baseline in the trials in which they said they had guessed, indicating unconscious knowledge of sentence processing. This effect was found in the NBD group across all argument orders and in the PWA group in sentences presented in A-T order. It suggests that PWA lack the required unconscious knowledge, or access to it, to guide the processing of sentences presented in T-A argument order.

Taken together, the data demonstrate that PWA show reduced metacognitive awareness of sentence processing, and hence little voluntary control over their assignment of thematic roles. Overall, there is no evidence that PWA use internal compensatory strategies to overcome their comprehension difficulties, as proposed for other linguistic domains; for example, word retrieval (see Tompkins et al., 2006) or conversational efficiency (Beeke et al., 2009). Thus, the data show that PWA do not systematically overuse an explicit (i.e., conscious) strategy such as agent-first heuristics to disentangle thematic roles to a higher degree than NBD, as the best probabilistic guess to reach the correct interpretation of the sentences. Indeed, the data show that PWA are mostly unaware of their comprehension failures.

### **4.5.3 Gaze fixation data**

Gaze fixation data were collected as a measure of online language processing to get insight into the parsing routines in Spanish speakers with aphasia and corresponding NBD participants. In line with previous findings (Arantzeta et al., 2017; Dickey et al., 2007; Hanne et al., 2011), the gaze data show that sentence comprehension accuracy in PWA was underlined by distinctive parsing routines. Correctly and incorrectly answered sentences corresponded to distinctive fixation patterns along the visual display. Gaze data analysis showed that the confidence rating provided by the participants trial-by-trial was a significant predictor of the gaze fixation pattern. PWA showed different gaze fixation proportions to the target and foil pictures in correctly and incorrectly judged trials. Trials answered by guessing consistently showed a midpoint in the proportion of fixations into the target/foil picture in relation to correctly and incorrectly judged trials.

The data suggest that there is a relationship between real time sentence processing and the subjective perception of correctness in sentence comprehension. Nonetheless, it is not possible to identify the causal direction of this relationship. That is, we do not know whether the gaze fixations have an effect on the subsequent individual perception of the correctness of the answer, or whether different parsing routines, corresponding to

different degrees of confidence, are reflected distinctively in the gaze data of PWA. So far, the data show that confidence ratings are a sensitive measure of metacognitive awareness.

#### 4.5.4 Error awareness and revision of parsing routines

So far, we have shown that PWA have little conscious knowledge of sentence processing, and probably, therefore, reduced voluntary control of thematic role assignment. For NBD both conscious and unconscious knowledge interplay in the parsing routines, as indicated by a moderate relationship between sentence accuracy and confidence ratings. It suggests that a degree of self-monitoring is crucial in sentence processing. These findings are compatible with the perceptual loop theory (Levelt, 1989; 2001), which states that the comprehension system regulates the internal speech as well as the self-produced overt speech. Thus, deficits in the perceptual feedback loop are responsible for the misdetection of speech errors. PWA analyzed in the current study were characterized by having agrammatic language. Thus, the perceptual loop theory would predict that the impaired feedback system would provide PWA with deficits in language comprehension monitoring.

Language monitoring is involved in repair and reanalysis when the linear Agent-Theme thematic role assignment (perhaps determined by an agent-first heuristic) does not guide the correct interpretation of the sentence. In PWA, the reduced metacognitive awareness (i.e., conscious knowledge) shown in the current study might be related to the inability to perceive the need for revision. Both the behavioral and the gaze data partially converge with the “deterministic parsing” suggested by Hanne et al. (2011), and support the view that PWA adopt either a canonical or a non-canonical sequence of thematic-role assignment from early on during the presentation of the sentence<sup>7</sup>. Nevertheless, contrary to Hanne et al. (2011), we have no evidence that PWA are occasionally aware of the need for revision, but fail to backtrack. Rather, we argue that PWA do not perceive the need to revise the initial parsing. We will defend this interpretation from two different perspectives.

First, PWA showed a tendency to look at either the target or foil picture from the presentation of the verb, while progressively increasing the proportion of fixations towards the selected picture. Contrary to Hanne et al. (2011), but in line with our previous findings in Basque speakers with aphasia (Chapter 2) and Spanish speakers with aphasia (Chapter 3), PWA did not shift fixations to the counterpart picture at any point of the

<sup>7</sup> This pattern fits with task-specific demands and their influence on sentence comprehension accuracy in PWA, as described by the across-task analysis of Caplan, Michaud, & Hufford (2013). As opposed to other tasks such as self-paced listening (Caplan & Waters, 2003; Caplan et al., 2013) or classical visual word paradigms (Dickey et al., 2007; Meyer et al., 2012) in sentence-picture matching tasks the scenes of the full representation of the oncoming linguistic stimuli are presented slightly before or at the same time as the linguistic stimuli. This may influence listeners to make an early commitment to interpret the sentence according to one or other visual representation, likely due to processing resource reduction (see Caplan et al., 2013).

presentation of the sentence. Based on the gaze fixation data, they did not show any attempt to revise the initial parsing routine.

Second, in healthy speakers, the psychological experience of revising the parsing routine is sometimes conducted by conscious problem-solving mechanisms (see Townsend & Bever, 2001). The heavy processing costs of certain constructions require increased conscious attention. If we consider the diminished cognitive processing abilities reported for PWA (Caplan et al., 2007; Kolk, 1993; Miyake et al., 1994), we expect that their conscious threshold in revising routines is significantly lower than in healthy controls. The need to revise parsing routines, and eventually their failure, might be available to conscious knowledge in sentence constructions that healthy speakers can revise unconsciously. PWA may feel the need to reanalyze when their initial parsing routine does not match with the presence/absence of the accusative preposition or verbal morphology. However, if they failed to revise, we expect them to have an intuitive judgment of processing difficulty (i.e., “not sure/guess”). On the contrary, PWA in our study seldom reported uncertainty in the sentence comprehension task, but incorrectly judged the incorrectly comprehended sentences (i.e., false negative detection). Altogether, these data suggest that PWA chose to parse the sentences following either linear or non-linear thematic role assignment essentially from the first thematic role determining component. They do not revise the parsing routine when it results in incorrect interpretation of the sentence. There is no evidence to suggest that PWA have a specific difficulty in reanalysis processes *per se*. Instead, data advocate that PWA do not perceive the need to revise the parsing routine in the absence of self-monitoring abilities.

A potential way to shed light on this debate would be to test the processing of fully ambiguous sentences presented auditorily, similar to the ones used by Erdocia et al. (2009) in Basque. Providing PWA with the option to match a fully ambiguous sentence with one and/or other visual representation of the stimuli would assess their ability to parse the sentence following a canonical or non-canonical template, as well as their ability to revise their initial parsing routine.

## 4.6 Conclusions

The study of metacognitive awareness, and hence consciousness has been widely neglected in aphasiology research regarding comprehension deficits. The current study introduced behavioral and online methods in consciousness research of PWA to get insight into the extent to which the participants' responses are mediated by conscious versus unconscious knowledge.

We aimed to investigate whether PWA consciously used compensatory strategies as agent-first heuristics, as the best probabilistic guess to overcome their limitation in the comprehension of semantically reversible sentences in non-canonical word orders. The study was conducted with a group of PWA and NBD speakers of Spanish; a language that allows a relatively flexible word order in sentential structures. Participants performed a sentence-picture matching task of sentences presented in linear (i.e., A-T) and derived (T-A) argument order, followed by a subjective rating of their awareness of the accuracy in the primary task. As expected, PWA comprehended significantly less well than NBD across all argument orders (i.e., A-T and T-A). Moreover, PWA were more impaired in their comprehension of sentences with a derived T-A argument order.

The relationship between comprehension accuracy and confidence rating was studied to get insight into the degree to which sentence processing was unconsciously mediated (i.e., zero-correlation criterion; Dienes et al., 1995). In the PWA group, the confidence rating was a very poor predictor of sentence comprehension accuracy, indicating reduced metacognitive awareness of comprehension failure. In contrast, confidence rating was a moderate-good predictor in the NBD group. Regarding the guessing-criterion (Dienes et al., 1995), PWA had above-chance performance in the comprehension of sentences presented in A-T order, even when they claimed to be guessing, but not in sentences presented in T-A order. This suggests that unlike NBD participants, PWA do not show signs of the implicit knowledge required for correct thematic role assignment. Note that participants believed themselves to be guessing in very few responses. On the one hand, this indicates that cautious interpretation of the guessing criterion is needed. On the other hand, the pattern of responses is not compatible with the predictions of the TDH (Grodzinsky, 1986, 1995, 2000; see Drai & Grodzinsky, 2006ab for a later revision), which claims that PWA answer by guessing when confronted with reversible sentences in non-canonical word orders. By replicating previous findings (Arantzeta et al., 2017, in Chapter 2; Hanne et al., 2011), gaze data analysis also contradicts this by showing that gaze fixation patterns diverge between correctly and incorrectly comprehended trials and the divergence starts at the first constituent to which a thematic role can be assigned. That is, PWA assign the role of either agent or theme to the first possible constituent, and (mostly, although not always) proceed on that basis.

In addition, gaze analysis showed that PWA look at the visual stimuli in a distinctive way depending on the confidence with which the trials were rated. Although the causal direction of this relationship cannot be determined in the current study, it demonstrates the validity of confidence rating for obtaining insight into unconscious processes in PWA. In summary, PWA showed anosognosic behavior towards sentence comprehension deficits, even when they were aware of their aphasic condition. These results converge with previous studies in jargon aphasia (see Rubens and Garret, 1991). The lack of conscious

cognition, which is strongly linked with voluntary control, does not suggest that they use compensatory strategies to overcome their comprehension difficulties. Thus, awareness cannot be taken as a self-regulatory mechanism for therapeutic applications in PWA, as suggested for healthy individuals (Koriat, 2000). We have speculated that the lack of conscious control might be congruent with the inhibition problems suggested by some authors to understand sentence comprehension deficits in PWA (Dickey et al., 2007; Hanne et al., 2011). Further research is required to determine the relationship between executive functions in relation to sentence comprehension deficits in PWA.